Decomposition and Identification of Linear Structural Equation Models

Bryant Chen

August 11, 2015

Abstract

In this paper, we address the problem of identifying linear structural equation models. We first extend the edge set half-trek criterion to cover a broader class of models. We then show that any semi-Markovian linear model can be recursively decomposed into simpler sub-models, resulting in improved identification power. Finally, we show that, unlike the existing methods developed for linear models, the resulting method subsumes the identification algorithm of non-parametric models.

1 Introduction

Many researchers, particularly in economics, psychology, and the social sciences, use linear structural equation models (SEMs) to describe the causal and statistical relationships between a set of variables, predict the effects of interventions and policies, and to estimate parameters of interest. When modeling using linear SEMs, researchers typically specify the causal structure (i.e. exclusion restrictions and independence restrictions between error terms) from domain knowledge, leaving the structural coefficients (representing the strength of the causal relationships) as free parameters to be estimated from data. If these coefficients are known, then total effects, direct effects, and counterfactuals can be computed from them directly (Pearl, 2009; Chen and Pearl, 2014). However, in some cases, the causal assumptions embedded in the model are not enough to uniquely determine one or more coefficients from the probability distribution, and therefore, cannot be estimated using data. In such cases, we say that the coefficient is *not identified* or *not identifiable*[1].

Many SEM researchers determine the identifiability of the model by submitting the specification and data to software, which attempts to estimate the coefficients by minimizing a fitting function (Kenny and Milan, 2012). If the model is not identified, then the program will be unable to complete the estimation and warns that the model may not be identified. While convenient, there are disadvantages to using typical SEM software to determine model identifiability (Kenny and Milan, 2012; Chen and Pearl, 2014). If poor starting values are chosen, the program could mistakenly conclude the model is not identified when in fact it may be identified. When the model is not identified, the program is not helpful in indicating which parameters are not identified nor are they able to provide estimates for identifiable coefficients[2]. Most importantly, the program only gives an answer after the researcher has taken the time to collect data.

Rather than determining the identifiability of parameters by attempting to fit the model, we will detect identifiability directly from the model specification and express identified parameters in terms of the population covariance matrix. As a result, the modeler can estimate their values, even when the model, as a whole, is not identified. Additionally, we avoid issues of poor starting values and can determine the identifiability of parameters prior to collecting data.

As compact and transparent representations of the model's structure, causal graphs provide a convenient tool to aid in the identification of coefficients. First utilized as a causal inference tool by Wright (1921), graphs have more recently been applied to identify causal effects in non-parametric causal models (Pearl,

---

[1]We will also use the term "identified" with respect to individual variables and the model as a whole. When the structural equation for a given variable contains one or more unidentified coefficients then we say the variable is unidentified. Similarly, when a model contains one or more unidentified coefficients than the model is not identified.

[2]According to Kenny and Milan (2012), AMOS is the only program that attempts to estimate parameters when the model is not identified.

2000) and enabled the development of causal effect identification algorithms that are complete for non-parametric models (Tian, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006). These algorithms can be applied to the identification of coefficients in linear SEMs by identifying non-parametric direct effects, which are closely related to structural coefficients (Tian, 2005; Chen and Pearl, 2014). Algorithms designed specifically for the identification of linear SEMs were developed by Brito and Pearl (2002), Brito (2004), Tian (2005, 2007, 2009), Foygel et al. (2012), and Chen et al. (2014).

Surprisingly, none of the aforementioned identification methods subsume the algorithms designed for non-parametric models when naively applied to the identification of linear SEMs[3], despite the added assumption of linearity. In this paper, we first extend the half-trek identification methods of Foygel et al. (2012) and Chen et al. (2014). Our extensions can be applied to both semi-Markovian and non-Markovian models. We then demonstrate how recursive c-component decomposition, which was first utilized in identification algorithms for semi-Markovian non-parametric models (Tian, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006), can be utilized for semi-Markovian linear models. We show that applying recursive decomposition to our identification method allows the identification of additional coefficients and models. Further, we will demonstrate that this method subsumes the aforementioned non-parametric identification methods when applied to linear SEMs. Lastly, we note that our method is not complete and there are identifiable coefficients in certain models that cannot be identified using the methods given.

2 Preliminaries

A linear structural equation model consists of a set of equations of the form,

$$X = \Lambda X + \epsilon,$$

where $X = [x_1, ..., x_n]^t$ is a vector containing the model variables, Λ is a matrix containing the *coefficients* of the model, which convey the strength of the causal relationships, and $\epsilon = [\epsilon_1, ..., \epsilon_n]^t$ is a vector of error terms, which represents omitted or latent variables. The matrix Λ contains zeroes on the diagonal, and $\Lambda_{ij} = 0$ whenever x_i is not a cause of x_j. The error terms are random variables and induce the probability distribution over the model variables. The covariance matrix of X will be denoted by Σ and the covariance matrix over the error terms, ϵ, by Ω.

An instantiation of a model M is an assignment of values to the model parameters (i.e. Λ and the non-zero elements of Ω). For a given instantiation m_i, let $\Sigma(m_i)$ denote the covariance matrix implied by the model and $\lambda_k(m_i)$ be the value of coefficient λ_j. In this paper, we identify the model coefficients from the covariance matrix, Σ.

Definition 1. *A coefficient, λ_k, is* identified *if $\lambda_k(m_i) = \lambda_k(m_j)$ whenever $\Sigma(m_i) = \Sigma(m_j)$.*

In other words, λ_k is identified if it can be uniquely determined from the covariance matrix, Σ.

The causal graph or path diagram of an SEM is a graph, $G = (V, D, B)$, where V are vertices or nodes, D directed edges, and B bidirected edges. The vertices represent model variables. Directed eges represent the direction of causality, and for each coefficient $\Lambda_{ij} \neq 0$, an edge is drawn from x_i to x_j. Each directed edge, therefore, is associated with a coefficient in the SEM, which we will often refer to as its structural coefficient. The error terms, ϵ_i, are not represented in the graph. However, a bidirected edge between two variables indicates that their corresponding error terms may be statistically dependent while the lack of a bidirected edge indicates that the error terms are independent. When the causal graph is acyclic without bidirected edges, then we say that the model is *Markovian*. Graphs with bidirected edges are *non-Markovian*, while acyclic graphs with bidirected edges are additionally called *semi-Markovian*.

If a directed edge, called (x, y), exists from x to y then we say that x is a parent of y. The set of parents of y is denoted $Pa(y)$. Additionally, we call y the head of (x, y) and x the tail. The set of tails for a set of directed edges, E, is denoted $Ta(E)$ while the set of heads is denoted $He(E)$. For a node, v, the set of edges for which $He(E) = v$ is denoted $Inc(v)$. If there exists a path of directed edges from x to y then we say that x is an ancestor of y. The set of ancestors of y is denoted $Anc(y)$. If x is an ancestor of y, then y is a descendant of x. The set of descendants of x is denoted $De(x)$. Finally, the set of nodes connected to y by a bidirected arc are called the siblings of y or $Sib(y)$.

---

[3]The coefficient f in Figure 2a is identified using non-parametric algorithms but not the aforementioned linear algorithms.

A *path* from x to y is a sequence of edges connecting the two vertices. A path may go either along or against the direction of the edges. A non-endpoint vertex w on a path is said to be a *collider* if the edges preceding and following w on the path both point to w, that is, $\rightarrow w \leftarrow$, $\leftrightarrow w \leftarrow$, $\rightarrow w \leftrightarrow$, or $\leftrightarrow w \leftrightarrow$. A vertex that is not a collider is a *non-collider*.

A path between x and y is said to be *unblocked given a set Z* (possibly empty), with $x, y \notin Z$ if

1. every noncollider on the path is not in Z and

2. every collider on the path is in $An(Z)$ (Pearl, 2009).

3 Extending the Edge Set Half-Trek Criterion

The half-trek criterion is a graphical condition that can be used to determine the identifiability of recursive and non-recursive linear models (Foygel et al., 2012). Foygel et al. (2012) use the half-trek criterion to identify the model variables one at a time, where each identified variable may be able to aid in the identification of other variables. If any variable is not identifiable using the half-trek criterion, then their algorithm returns that the model is not *HTC-identifiable*. Otherwise the algorithm returns that the model is identifiable. Their algorithm subsumes the earlier methods of Brito and Pearl (2002) and Brito (2004). Chen et al. (2014) modified the half-trek criterion (calling it the *edge set half-trek criterion*) to identify *connected edge sets*. Rather than attempting to identify all of a variable's coefficients at once, they instead partition the coefficients according to connected edge sets. As a result, if one coefficient is not identified, only its connected edge set will not be identified rather than the entire variable. In this way, they increase the granularity of the criterion to identify additional coefficients in unidentifiable models. In this section, we will paraphrase some preliminary definitions from (Foygel et al., 2012) and further extend the edge set half-trek criterion. Without c-component decomposition, this extension will allow identification of even more coefficients in unidentifiable models. Coupled with recursive c-component decomposition, our extended half-trek criterion allows the identification of additional models, as well.

3.1 Preliminary Definitions

We begin by establishing a couple preliminary definitions around half-treks. These definitions and illustrative examples can also be found in Foygel et al. (2012) and Chen et al. (2014).

Definition 2. *(Foygel et al., 2012) A* half-trek, *π, from x to y is a path from x to y that either begins with a bidirected arc and then continues with directed edges towards y or is simply a directed path from x to y.*

We will denote the set of nodes connected to a node, v, via half-treks $htr(v)$.

Definition 3. *(Foygel et al., 2012) For any half-trek, π, let* Right(π) *be the set of vertices in π that have an outgoing directed edge in π (as opposed to bidirected edge) union the last vertex in the trek. In other words, if the trek is a directed path then every vertex in the path is a member of Right(π). If the trek begins with a bidirected edge then every vertex other than the first vertex is a member of Right(π).*

Definition 4. *(Foygel et al., 2012) A system of half-treks, $\pi_1, ..., \pi_n$, has* no sided intersection *if for all $\pi_i, \pi_j \in \{\pi_1, ..., \pi_n\}$ such that $\pi_i \neq \pi_j$, Right$(\pi_i) \cap$ Right$(\pi_j) = \emptyset$.*

Definition 5. *(Chen et al., 2014) For an arbitrary variable, V, let $Pa_1, Pa_2, ..., Pa_k$ be the unique partition of $Pa(V)$ such that any two parents are placed in the same subset, Pa_i, whenever they are connected by an unblocked path. A* connected edge set *with head V is a set of directed edges from Pa_i to V for some $i \in \{1, 2, ..., k\}$.*

3.2 General Half-Trek Criterion

Having established the preliminary definitions in the previous subsection, we now give our extension of the edge set half-trek criterion.

Definition 6. *(General Half-Trek Criterion) Let E be a set of directed edges sharing a single head v. A set of variables Y satisfies the* general half-trek criterion *with respect to E, if*

(i) $|Y| = |E|$,

(ii) $Y \cap (v \cup Sib(v)) = \emptyset$,

(iii) *There is a system of half-treks with no sided intersection from Y to $Ta(E)$, and*

(iv) $(Pa(v) \setminus Ta(E)) \cap htr(Y) = \emptyset$.

A set of directed edges, E, sharing a head v is identifiable if there exists a set, Y_E, that satisfies the general half-trek criterion (g-HTC) with respect to E, and Y_E consists only of "allowed" nodes. Intuitively, a node, y, is not allowed if

(i) it is half-trek reachable from $He(E)$ and the coefficients of y that lie on the half-treks from $He(E)$ are not identifiable or

(ii) y is connected to $Pa(v) \setminus Ta(E)$, and the coefficients of y that lie on unblocked paths between $Pa(v) \setminus Ta(E)$ are not identifiable.

Otherwise, y is allowed. The following definition formalizes this notion.

Definition 7. *Let E be the set of directed edges in a causal graph, G. We say that a node, z, is g-HT-allowed (or simply allowed) for directed edges $E_v \subseteq E$ with head v if*

(i) (a) *z is not half-trek reachable from v and*

(b) *z is not connected to $Pa(v) \setminus Ta(E)$*

OR

(ii) *there exists a partition of E, ES, such that $E_v \in ES$, an ordering on ES, $\prec$, and a family of subsets (Y_{E_i}), one subset for each $E_i \prec E_v$, such that Y_{E_i} satisfies the g-HTC with respect to E_i and $E_j \prec E_k$ for $(E_j, E_k \prec E_v)$ whenever*

(a) $He(E_j) \subseteq htr(He(E_k)) \cap Y_{E_k}$ *or*

(b) *Y_{E_k} is connected to $Pa(He(E_k)) \setminus Ta(E_k)$ via edges in the set $Inc(Y_{E_k}) \cap E_j$*

and any directed edges belonging to z that lie on a half-trek from v to z or lie on a path between z and $Pa(He(E_k)) \setminus Ta(E_k)$ belong to a set $E_z \subset ES$ ordered before E_v.

If Y is a set of allowed variables for E_v that satisfies the half-trek criterion with respect to E_v, we will say that Y is an *g-HT-admissible* set for E_v. We are now ready to use the g-HTC to identify coefficients.

Theorem 1. *If a g-HT-admissible set for directed edges E_v with head v exists then E_v is identifiable. Further, let $Y_{E_v} = \{y_1, ..., y_k\}$ be a g-HT-admissible set for E_v, $Ta(E_v) = \{p_1, ..., p_k\}$, and Σ be the covariance matrix of the model variables. Define $\mathbf{A}$ as*

$$\mathbf{A_{ij}} = \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i p_j}, & y_i \in htr(v) \text{ or } y_i \text{ connected} \\ & \text{to } Pa(v) \setminus Ta(E_v), \\ \Sigma_{y_i p_j}, & y_i \notin htr(v) \end{cases} \tag{1}$$

and $\mathbf{b}$ as

$$\mathbf{b_i} = \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i v}, & y_i \in htr(v) \text{ or } y_i \text{ connected} \\ & \text{to } Pa(v) \setminus Ta(E_v), \\ \Sigma_{y_i v}, & y_i \notin htr(v) \end{cases} \tag{2}$$

Then $\mathbf{A}$ is an invertible matrix and $\mathbf{A} \cdot \Lambda_{Ta(E_v), V} = \mathbf{b}$.
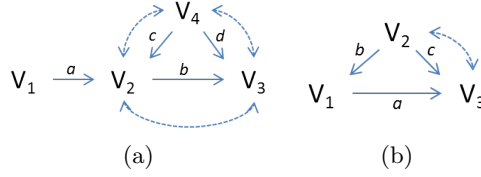
4

Figure 1: (a) b is identified using the g-HTC but not the edge set HTC (b) a is identified using the g-HTC but not the edge set HTC

Proof. The proof for this theorem is similar to the proof of Theorem 1 in Foygel et al. (2012). Rather than giving a complete proof, we simply explain why our changes are valid. The g-HTC identifies arbitrary sets of directed edges belonging to a node rather than all of the directed edges belonging to a node. It is able to do this because of two changes. First, sets that contain nodes that are connected to $Pa(v) \setminus Ta(E)$ via half-treks cannot be half-trek admissible for E (see Definition 6). As a result, the paths from half-trek admissible set, Y_E, to v travel only through coefficients of E and no other coefficients of E. This ensures that $\mathbf{A} \cdot \Lambda_{Ta(E),v} = \mathbf{b}$. Second, nodes that are connected to $Pa(v) \setminus Ta(E)$ are not allowed unless their coefficients that lie on paths to $Pa(v) \setminus Ta(E)$ are identified. Likewise, nodes that are half-trek reachable from v are not allowed unless their coefficients that lie on the half-treks from v are identified. This ensures that $\mathbf{A}$ and $\mathbf{b}$ are computable. Other coefficients need not be identified because they will vanish from $\mathbf{A}$ and $\mathbf{b}$ during the computations, $((I - \Lambda)^T \cdot \Sigma)_{y_i p_j}$ and $((I - \Lambda)^T \cdot \Sigma)_{y_i v}$, due to zeroes in the matrix Σ. □

The g-HTC impoves upon the edge set HTC because subsets of connected edge sets may be identifiable even when the connected edge set as a whole is not.

Theorem 2. *Any coefficient identifiable using the edge set HTC is also identifiable using the g-HTC. Moreover, the g-HTC is able to identify coefficients in certain models that the edge set HTC is not.*

Proof. First, we note that when the set of edges in question is a connected edge set, then the edge set HTC and the g-HTC are equivalent. As a result, any set of edges identifiable using the edge set HTC is identifiable using the g-HTC, and we have proven the first part of the Theorem.

We will now use Figures 1a and 1b to show that certain coefficients that are not identifiable using the edge set HTC are identifiable using the g-HTC. The edge set HTC identifies coefficients associated with connected edge sets altogether. As a result, the coefficient b in the graph depicted in Figure 1a is not identifiable using the edge set HTC since it belongs to the connected edge set $\{V_2 \to V_3, V_4 \to V_3\}$ and the coefficient d is not identifiable. However, b is identifiable using the g-HTC since V_1 is g-HT-admissible for b.

Additionally, coefficient a in Figure 1b is not identifiable using the edge set half-trek criterion due to the fact that c is not identifiable. Using the g-HTC and Theorem 1, we can first identify b since V_2 is a g-HT-admissible set for b. Once b is identified, we can use $Y_a = V_1$ to identify a. □

3.3   g-HTC Algorithm

Finding a g-HT-admissible set for directed edges, E, with head, v, from a set of allowed nodes, A_E, can be accomplished by utilizing the max-flow algorithm described in Chen et al. (2014)[4], which we call MaxFlow($G_f(E, A)$). This algorithm returns a maximal set of allowed nodes that satisfies (ii) - (iv) of the g-HTC.

In some cases, there may be no g-HT-admissible set for E' but there may be one for $E \subset E'$. In other cases, there may be no g-HT admissible set of variables for a set of edges E but there may be a g-HT admissible set of variables for E' with $E \subset E'$. As a result, if a HT-admissible set does not exist for E_v, where $E_v = Inc(v)$ for some node v, we may have to check whether such a set exists for all possible subsets of E_v in order to identify as many coefficients in E_v as possible. This process can be somewhat simplified

---

[4]Brito (2004) utilized a similar max-flow construction in his identification algorithm. This was later modified by Foygel et al. (2012) for the half-trek criterion and finally Chen et al. (2014) for the edge set half-trek criterion.
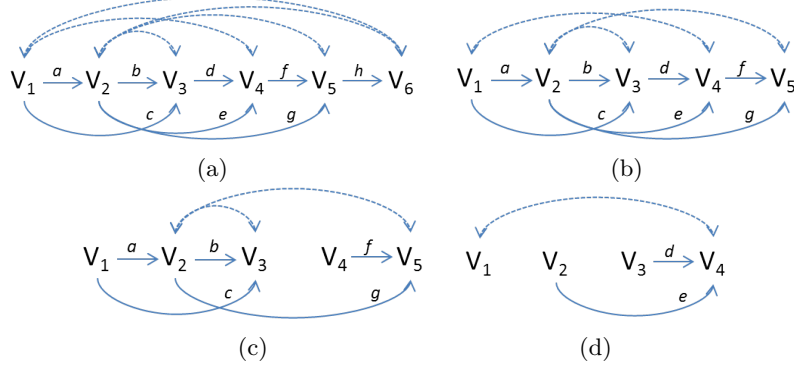
Figure 2: (a) The graph is not identified using the g-HTC and cannot be decomposed (b) After removing V_6 we are able to decompose the graph (c) Graph for c-component, $\{V_2, V_3, V_5\}$ (d) Graph for c-component, $\{V_1, V_4\}$

by noting that if E is a connected edge set with no g-HT-admissible set, then there is no superset E' with a g-HT-admissible set.

Algorithm 1 utilizes the g-HTC and Theorem 1 to identify as many coefficients in a recursive or non-recursive linear SEM as possible. It iterates through every unidentified connected edge set in the graph, attempting to identify each unidentified subset of its edges using Theorem 1. For a given set of directed edges, E, the initial set of HT-allowed nodes is the set of nodes that is not half-trek reachable from $He(E)$. As more and more edges are identified, the set of allowed nodes increases so that a set of edges that was not initially identified may become identifiable as the algorithm iterates. We define Allowed(E, IDEdges, G) to be the set of nodes that are either (i) not half-trek reachable from $He(E)$ and not connected to $Pa(He(E)) \setminus Ta(E)$ or (ii) their edges that lie on half-treks from $He(E)$ or paths to $Pa(He(E)) \setminus Ta(E)$ are in IDEdges. As a result, Algorithm 1 attempts to identify each connected edge set (and its subsets) until all coefficients have been identified or no new coefficients are identified in a given iteration.

Algorithm 1 HT-ID(G, Σ, IDEdges)

    Initialize: EdgeSets $\leftarrow$ all connected edge sets in G
    repeat
        for each ES in EdgeSets **do**
            $v \leftarrow He(ES)$
            for each $E \subset ES$ such that $E \not\subset$ IDEdges **do**
                $A_E \leftarrow$ Allowed(E, IDEdges, G)
                $Y_E \leftarrow$ MaxFlow($G_f(v, A_E)$)
                if $|Y_E| = |Ta(E)|$ **then**
                    Identify E using Theorem 1
                    IDEdges $\leftarrow$ IDEdges $\cup E$
                end if
            end for
        end for
    until All coefficients have been identified or no coefficients have been identified in the last iteration
    return IDEdges

---

4 Decomposition

Tian (2005) showed that the identification problem could be simplified in semi-Markovian linear structural equation models by decomposing the model into sub-models according to their c-components, defined below. Each coefficient is identifiable if and only if it is identifiable in the sub-model to which it belongs. Foygel et al. (2012) noted that first decomposing the model in this way and then applying the half-trek criterion

allows for the identification of a larger set of models. In this section, we will briefly review the results of Tian (2005) before showing that the c-component decomposition can be applied recursively to the model after marginalizing certain variables. This idea was first used to identify interventional distributions in non-parameteric models by Tian (2002) and Tian and Pearl (2002) and adapting this technique for linear models will allow us to identify models that the g-HTC, even coupled with (non-recursive) c-component decomposition, is unable to identify. Further, it ensures the identification of all coefficients identifiable using methods developed for non-parametric models–a guarantee that none of the existing methods developed for linear models satisfy.

4.1   C-Component Decomposition

Definition 8. *(Tian, 2005) The* c-component *of a node v in a causal graph, denoted $C(v)$, is the set of nodes that is connected to v by paths consisting only of bidirected arcs.*

Tian (2005) showed that a coefficient is identified if and only if it is identified in the sub-graph consisting of its c-component and the parents of the c-component. More formally, a coefficient from x to y is identified if and only if it is identified in the sub-model constructed in the following way:

(i) The sub-model variables consist of $C(y) \cup Pa(C(y))$.

(ii) The structural equations for the variables in $C(y)$ are the same as their structural equations in the original model while the structural equations for the parents simply equate each parent with its error term.

(iii) If the error terms of any two variables in the sub-model were uncorrelated in the original model then they are uncorrelated in the sub-model.

When we decompose the model according to c-components, the distribution over the variables in the sub-model, $W = S \cup Pa(S)$, for a c-component S, is not simply the marginal distribution over W. Instead, the distribution, which we denote P_S, is computed from the joint distribution over the model variables, $P(V)$. Letting $V = \{v_1, v_2, ..., v_n\}$ be ordered topologically, we have

$$P_S(W) = \prod_{\{i|v_i \in S\}} P(v_i|V^{(i-1)}) \prod_{\{l|v_l \in Pa(S)\}} P(v_l), \tag{3}$$

where $V^{(i-1)} = \{v_1, v_2, ..., v_{i-1}\}$.

For example, consider the graph depicted in Figure 2b. We have two c-components, $S_1 = \{v_2, v_3, v_5\}$ and $S_2 = \{v_1, v_4\}$, with

$$P_{S_1} = P(v_2|v_1)P(v_3|v_2, v_1)P(v_5|v_4, v_3, v_2, v_1)$$
$$P(v_1)P(v_4)$$

and

$$P_{S_2} = P(v_1)P(v_4|v_3, v_2, v_1)P(v_3)P(v_2)P(v_3).$$

We will denote the sub-model for a c-component, S, M_S and the graph it induces G_S. G_{S_1} is depicted in Figure 2c and G_{S_2} in Figure 2d.

Theorem 3. *(Tian, 2005) Let a variable v_j be in a c-component S_j in a SEM M. A coefficient, Λ_{ij}, is identifiable if and only if it is identifiable in the model M_{S_j}.*

4.2   Recursive C-Component Decomposition

The graph in Figure 2a consists of a single c-component, and we are unable to decompose it. As a result, we are able to identify a but no other coefficients using the extended half-trek criterion. Moreover, $f = \frac{\partial}{\partial v_4} E[v_5|do(v_6, v_4, v_3, v_2, v_1)]$ is identified using identification methods developed for non-parametric models (e.g. do-calculus) but not the g-HTC or other methods developed for linear models.

7

However, if we remove v_6 from the analysis, then the resulting model can be decomposed. Let M be the model depicted in Figure 2a, $P(v)$ be the distribution induced by M, and M' be a model that is identical to M except the equation for v_6 is removed. M' induces the distribution $\int_{v_6} P(V) dv_6$, and its associated graph G' yields two c-components, as shown in Figure 2b.

Now, decomposing G' according to these c-components yields the sub-models depicted by Figures 2c and 2d. Both of these sub-models are identifiable using the half-trek criterion. Thus, all coefficients other than h have been shown to be identifiable. Returning to the graph prior to removal, depicted in Figure 2a, we are now able to identify h because both v_4 and v_5 are now allowed nodes for h, and the model is identified[5].

As a result, we can improve our identification algorithm by recursively decomposing, using the extended half-trek criterion, and removing descendant sets[6]. Note, however, that we must consider every descendant set for removal. It is possible that removing D_1 will allow identification of a coefficient but removing a superset D_2 with $D_1 \subset D_2$ will not. Additionally, it is possible that removing D_2 will allow identification but removing a subset D_1 will not.

After recursively decomposing the graph, if some of the removed variables were unidentified, we may be able to identify them by returning to the original graph prior to removal since we may have a larger set of allowed nodes. For example, we were able to identify h in Figure 2a by "un-removing" v_6 after the other coefficients were identified. In some cases, however, we may need to again recursively decompose and remove descendant sets. As a result, in order to fully exploit the powers of decomposition and the g-HTC, we must repeat the recursive decomposition process on the original model until all marginalized nodes are identified or no new coefficients are identified in an iteration.

Algorithm 2 decomposes the graph according to its c-components and then applies HT-ID (Algorithm 1) to each sub-model. If there are still unidentified coefficients, then it removes descendant sets and decomposes again. The whole process is repeated until every coefficient is identified or no new coefficients are identified in an iteration. $\Sigma_{P_{S_i}}$ is the covariance matrix of P_{S_i}, where S_i is a c-component. $\Sigma_{V \setminus D_i}$ is the covariance matrix after marginalizing D_i from Σ. Finally, $G_{V \setminus D_i}$ is the graph with the set D_i removed.

Algorithm 2 Decomp-HT-ID(G, Σ)

Initialize: IDEdges $\leftarrow \emptyset$
repeat
    IDEdges $\leftarrow$ IDEdges$\cup$Rec-Decomp$(G, \Sigma, \text{IDEdges})$
until All coefficients have been identified or no coefficients have been identified in the last iteration
return IDEdges

procedure Rec-Decomp$(G, \Sigma, \text{IDEdges}))$
    $V \leftarrow$ vertices in G
    Edges $\leftarrow$ all edges in G
    for each c-component, S_i, in G **do**
        IDEdges $=$ IDEdges$\cup$ HT-ID$(G_{S_i}, \Sigma_{S_i}, \text{IDEdges})$
    end for
    if IDEdges $=$ Edges **then**
        Return IDEdges
    else
        for each descendant set, D_i, in G **do**
            IDEdges $\leftarrow$ IDEdges$\cup$Rec-Decomp$(G_{V \setminus D_i}, \Sigma_{V \setminus D_i}, \text{IDEdges}$
        end for
    end if
    return IDEdges
end procedure

---

[5]While v_4 and v_5 are technically not allowed according to Definition 7, they can be used in g-HT-admissible sets to identify h using Theorem 1 since their coefficients have been identified.

[6]Only removing descendant sets have the ability to break up -components. For example, removing $\{v_2\}$ from Figure 2a does not break the c-component because removing v_2 would relegate its influence to the error term of its child, v_3. As a result, the graph of the resulting model would include a bidirected arc between v_3 and v_6, and we would still have a single c-component.

We now show that any direct effect identifiable using non-parametric methods is also identified using Algorithm 2.

Theorem 4. *Let M be a linear SEM with variables V. Let $M^{'}$ be a non-parametric SEM with identical structure to M. If the direct effect of x on x for $x, y \in V$ is identified in $M^{'}$ then the coefficient Λ_{xy} in M is identified using Algorithm 2.*

Proof. Let G be the causal graph of M and $M^{'}$. Suppose the direct effect of x on y is identified in $M^{'}$. Then according to Theorem 3 of (Shpitser, 2008), there does not exist a subgraph of G that is a y-rooted c-tree (Shpitser, 2008). This implies that $MACS(y) = y$. By recursively decomposing the graph into c-components and marginalizing descendant sets, we can obtain a graph where only $MACS(y)$ and its parents remain in the graph. Since $MACS(Y) = y$, the parents of y in this graph represent a g-HT admissible set that allows the identification of all coefficients of y. □

5 Conclusion

In this paper, we extend the edge set half-trek criterion (Chen et al., 2014). We then incorporate recursive c-component decomposition (Tian, 2002), and show that the resulting identification method is able to identify more coefficients and models than the existing algorithms developed for both linear and non-parametric models.

6 Acknowledgments

References

BRITO, C. (2004). *Graphical methods for identification in structural equation models.* Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.
URL $<$http://ftp.cs.ucla.edu/pub/stat_ser/r314.pdf$>$

BRITO, C. and PEARL, J. (2002). Generalized instrumental variables. In *Uncertainty in Artificial Intelligence, Proceedings of the Eighteenth Conference* (A. Darwiche and N. Friedman, eds.). Morgan Kaufmann, San Francisco, 85–93.

CHEN, B. and PEARL, J. (2014). Graphical tools for linear structural equation modeling. Tech. Rep. R-432, <http://ftp.cs.ucla.edu/pub/stat_ser/r432.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Forthcoming, Psychometrika.

CHEN, B., TIAN, J. and PEARL, J. (2014). Testable implications of linear structual equation models. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence* (C. E. Brodley and P. Stone, eds.). AAAI Press, Palo, CA. <http://ftp.cs.ucla.edu/pub/stat_ser/r428-reprint.pdf>.

FOYGEL, R., DRAISMA, J. and DRTON, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics* **40** 1682–1713.

HUANG, Y. and VALTORTA, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 217–224.

KENNY, D. A. and MILAN, S. (2012). Identification: A non-technical discussion of a technical issue. In *Handbook of Structural Equation Modeling* (G. M. S. W. R. Hoyle, D. Kaplan, ed.). Guilford Press, New York, 145–163.

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York. 2nd edition, 2009.

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge University Press, New York.

SHPITSER, I. (2008). *Complete Identification Methods for Causal Inference.* Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.

SHPITSER, I. and PEARL, J. (2006). Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 437–444.

TIAN, J. (2002). *Studies in Causal Reasoning and Learning.* Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.

TIAN, J. (2005). Identifying direct causal effects in linear models. In *Proceedings of the National Conference on Artificial Intelligence*, vol. 20. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

TIAN, J. (2007). A criterion for parameter identification in structural equation models. In *Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07).* AUAI Press, Corvallis, Oregon.

TIAN, J. (2009). Parameter identification in a class of linear structural equation models. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09).*

TIAN, J. and PEARL, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence.* AAAI Press/The MIT Press, Menlo Park, CA, 567–573.

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20** 557–585.